DOCUMENT RESUME

ED 160 633 TM 007 713

AUTHOR Rakow, Ernest A.
TITLE Ridge Regression: A Regression Procedure for Analyzing Correlated Independent Variables.
PUB DATE Mar 78
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Academic Achievement; *Correlation; Elementary Secondary Education; Grade 4; Grade 7; *Multiple Regression Analysis; Predictive Validity; *Predictor Variables; Suppressor Variables
IDENTIFIERS *Ridge Regression Analysis

ABSTRACT

Ridge regression is presented as an analytic technique to be used when predictor variables in a multiple linear regression situation are highly correlated, a situation which may result in unstable regression coefficients and difficulties in interpretation. Ridge regression avoids the problem of selection of variables that may occur in stepwise procedures. Ridge regression is performed by adding a constant to the diagonal of the correlation matrix. This may be done iteratively for a range of constants, say from 0 to 3 in small increments. The results of data analysis are presented both in tables and in graphs, to show how ridge regression reduces the instability of regression weights in cross validdation. These data were generated to predict fourth and seventh grade achievement from achievement test scores and teacher variables. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED160633

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

RIDGE REGRESSION: A REGRESSION PROCEDURE FOR ANALYZING CORRELATED INDEPENDENT VARIABLES

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Ernest A. Rakow

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Ernest A. Rakow

Memphis State University

A paper presented at the Annual Meeting of the American Educational Research Association, Toronto, Ontario, Canada, March 27-31, 1978.

TM007 713

AERA Session 3.10

RIDGE REGRESSION: A REGRESSION PROCEDURE FOR ANALYZING CORRELATED INDEPENDENT VARIABLES

Ernest A. Rakow

Memphis State University

Multiple linear regression is frequently applied in educational research, especially in large scale survey type studies (e.g. Coleman, 1966; Husen, 1967; Peaker, 1975; Madaus, Kellaghan and Rakow, 1975). A difficult problem encountered in such studies has been caused by multicolinearity of independent variables. This multicolinearity creates difficulties in deciding which variables to include in the regression equation, in estimation of the regression coefficients and in the interpretation of relative importance of the variables via these regression coefficients.

The purpose of this paper is to present ridge regression. Ridge regression is an analytic technique to be utilized when predictor variables in the multiple linear regression situation are highly correlated. Ridge regression may predict nearly as accurately as multiple regression as well as yielding regression coefficients which are more easily interpreted.

Multiple linear regression with highly correlated predictor variables has several pitfalls. One is that there may be a suppressor variable for which the regression coefficient and correlation with the criterion may have opposite signs. Interpretation of the importance of a suppressor variable to the equation is very difficult. A closely related pitfall is when two variables are highly correlated and one has a large regression coefficient while the other has a very small coefficient. These differences in magnitude may have little meaning in terms of

the research problem. Still another problem with multicolinearity is that the regression coefficients may be unstable, i.e. slight changes in the data may cause noticeable changes in the regression coefficients. A final difficulty in linear regression with highly correlated predictors is that although the errors of estimation have a small variance in the estimation sample, this variance generally is much larger on a cross-validation sample.

In classical multiple linear regression, a solution to some of these pitfalls has been to follow a step-wise selection procedure. Then, in addition to interpreting the regression coefficients, attention is focused upon the multiple correlation and upon changes in the multiple correlation squared as additional variables are added into the equation. In this type of interpretation, the variable entered first (either by plan or analytic selection) is interpreted as having exaggerated importance since it is credited with contributing its own unique portion and that shared with all other variables entered later. Variables entered into the regression equation late in the process are given less importance because the portion they share with other variables is already included as part of those variables. To deal with this difficulty, Newton and Spurrell (1967) developed a system for identifying the proportion of unique and shared explained variances. Another major weakness with step-wise regression is that variables not selected for inclusion are assigned regression coefficients of zero while allowing the coefficients for variables selected to be larger than they should be. In such situations the resulting regression coefficients may be very difficult to interprete in that research situation.

Ridge regression analysis (as developed by Draper, 1963; Hoerl, 1962; Hoerl and Kennard, 1970a, 1970b and first presented to educators by Williams, 1975) presents an alternative procedure. It does not select variables, but rather utilizes all variables in the model. This places the burden of variable selection upon the researcher and his development of an adequate theoretical model. For a pair of highly correlated predictor variables, ridge regression would use a portion of each in the predictions. In effect, this procedure equalizes the predictive effects of highly correlated variables. Consequently, this procedure rejects no variables and tends to have no variables with opposite signs for the regression coefficient and correlation.

Ridge regression is performed by adding a constant to the diagonal of the correlation matrix and then calculating the standardized regression coefficients and multiple correlation squared. This may be done iteratively by adding the constant and generating the regression coefficients for a range of constants. The addition of a small positive number to the diagonal functions to reduce the effect of the multicolinearity in the predictor variables. Marquardt and Snee (1975) point out that the lowest constant with stable regression coefficients identifies the best ridge solution. A low constant is desired because as the constant is increased the multiple correlation decreases (i.e. the standard error of estimate increases). Different values from 0.0 to 3.0, in small increments, may be tried as the constant. When zero is used as the constant, the result is identical to multiple linear regression.

Some have pointed out that the regression coefficients in ridge regression are biased estimates but have smaller standard errors than classical regression

(Marquardt and Snee, 1975 and Hoerl, Kennard and Baldwin, (1975)). The gain in reduction of error for these coefficients more than compensates for the loss introduced by the bias. The smaller standard errors indicate the regression coefficients are more stable.

The data analyzed to demonstrate an application of ridge regression in this paper are school district means from a statewide assessment program. The 566 school districts with both fourth and seventh grades were systematically divided into two samples by placing every second district into Sample A and the others into Sample B. The same analyses were run on both samples to enable examination of the stability of results from one sample to another and to allow for cross-validation. The first example utilizes the fourth grade district means on four achievement subtests to predict seventh grade Composite Achievement. The second example utilizes five teacher variables to predict seventh grade Composite Achievement. The third example utilizes the nine predictors from the two prior examples (four achievement subtests and five teacher variables) to again predict seventh grade Composite Achievement. The high degree of multicolinearity between the achievement subtests and between the teacher variables would lead to unsatisfactory results with classical multiple linear regression because the regression coefficients would be uninterpretable. These examples demonstrate how this difficulty is met by ridge regression analysis.

In the first example, school district mean level of achievement as measured at grade seven by Composite Achievement is predicted by fourth grade mean levels of achievement as measured by tests of Word Relationships, Reading, English and Mathematics. The correlations among these predictor variables range from a low

of .80 for Mathematics and Reading in Sample A to a high of .90 for Word Relationships and Reading in both Samples A and B. These predictor variables are highly correlated.

Ridge regression analysis analyzes the correlation matrix to obtain estimates of the squared multiple correlation and standardized partial regression coefficients. In the analysis for this paper, the same constant is added to all diagonal elements in the correlation matrix and a regression equation is determined. The values used for this constant ranged from 0.00 to 0.20 in increments of 0.02, from .20 to 1.00 in increments of 0.05 and from 1.00 to 2.50 in increments of 0.10. The many small values are utilized because that is when the estimates show the greatest fluctuation. It should be noted that when the constant is 0.00 a classical multiple regression is obtained.

Figure 1 presents the results of this ridge regression analysis for Sample A. The lines illustrate the changes in the standardized regression coefficients for a variable in relation to changes in the constant added to the diagonal. Since none of the lines cross, the variables remain in the same order of importance. However, the magnitude of some beta weights changes dramatically. The initial beta for Reading is 0.257 when the constant k = 0, but it decreases rapidly to .180 when k = .20 and to .144 when k = .50. Word Relationships also has a large beta initially (.237) which also decreases rapidly to .139 when k = .50. In contrast, the beta for English begins negative (-.069), changing to positive (.003) when k = .08 and reaches a maximum of .077 when k = .90. The line for Mathematics is not similar to any of the others. Its first beta is .079, increases to .090 when k=.50 and then tails of slightly. Examination of this figure leads to the use of k = .50 as perhaps best. From this point the lines are nearly straight.

Since the multiple correlation squared tends to decrease as the constant increases, the lowest value of k associated with the straight portions of the lines is best.

Figure 2 presents the corresponding ridge regression analysis for Sample B. For Sample B, Mathematics has a large initial beta (.409). Beta for Mathematics decreases rapidly to .315 for k = .10, .269 for k = .20 and .205 for k = .50. Betas for Word Relationships and English begin at approximately .04 and increase to .11 or .12. Beta for reading drops initially, but stabilizes quickly. Note that in both Figure 1 and Figure 2, when k = 0 some variables have large regression weights while others are small. The effect of ridge regression is to reduce this variability in the weights of correlated variables. Table 1 presents the multiple correlation squared and standardized regression coefficients for three arbitrarily selected values of k.

In the second example, school district mean level of Composite Achievement is predicted by five teacher variables. They are (1) professional instructional staff per 1000 students (INSTRS/K), (2) Teachers per 1000 students (TCHRS/K), (3) years of teaching experience (TCHEXPER), (4) teacher's annual salary (TCHSALRY) and (5) percent of teachers with a masters degree (%MA). These variables also are school district means. The correlations among these predictor variables range from −.23 for INSTRS/K with TCHSALRY to a high of .97 for INSTRS/K with TCHRS/K. Both extreme correlations happen to occur in Sample A.

Figure 3 presents the Sample A results of these teacher variables being used as predictors of grade seven achievement. The betas for INSTRS/K and TCHRS/K are of particular interest. For multiple regression their betas are −.675 and .797 respectively. Under ridge regression however, these estimates change rapidly so that when k = .20 these betas have become −.035 and .160. When k = .40 the beta for INSTRS/K becomes positive. The betas for %MA and TCHSALRY also change considerably. When the constant reaches .50 these betas seem to have stabilized.

After this point, the lines all appear relatively straight. Also, the betas for °/oMA and TCHRS/K are nearly equal for all large values of the constant.

Figure 4 presents the corresponding Sample B results. These results are very similar to those for Sample A. Again, INSTRS/K and TCHRS/K change very rapidly for small values of the constant. However, for Sample B the squared multiple correlation is small, initially being .050 and then decreasing.

For the third example the four district mean levels of achievement at grade four are combined with the five district means of teacher variables for prediction of composite achievement at grade 7. These results are in Figure 5, Figure 6 and Table 1.

Figure 5 presents the standardized regression coefficients derived from Sample A. The shape of the line for each variable's beta is very similar to that of each variable in Figure 1 or Figure 3. The major exception is Mathematics which in Figure 5 begins with a beta of −.096. Word Relationships begins with a higher beta of .413 as compared to .237 in Figure 1. Note that after the betas stabilize (about k = .80) some lines are nearly coincident because the betas are so nearly equal, e.g. Reading with Word Relationships, English with Mathematics and with TCHEXPER, and INSTRS/K with °/oMA. These nearly equal standardized regression coefficients would seem to have intuitive meaning. It is counter instructive for district mean levels of fourth grade achievement to have negative weights for prediction of seventh grade achievement. Following ridge regression these variables would have positive betas. Also, the extremely different betas for INSTRS/K and TCHRS/K appear meaningless for two variables so similar in construction. Ridge regression provides more similar, but still unequal, weights for them.

Figure 6 presents the standardized regression coefficients for the same nine variables but from Sample B. Once again the shape of the line for each variable's

betas is very similar to that of the same variable in Figure 2 or Figure 4. The beta for English is initially much lower, but quickly reaches its level of about .11. Word Relationships begins near its desirable level of .12. Again Reading, English and Word Relationships end with nearly coincident lines. The teacher variables INSTRS/K and TCHRS/K appear to have rather different patterns of betas than in Figure 4. Specifically they are much more similar to each other and TCHRS/K starts out below INSTRS/K where previously it began with very large betas.

The beta weights from ridge regression with k = .50 (or even .80) appear to be more easily interpretable on an intuitive level than those from multiple regression. The betas from two different samples are also more similar when k = .50. The price paid for this increase in interpretability appears to be smaller squared multiple correlations. However, according to Hoerl, Kennard and Baldwin (1975) they should more accurately reflect the findings in a cross-validation.

A cross-validation study was done with this data. Every set of standardized regression coefficients calculated via the ridge regression analysis was applied to the standardized scores from both Sample A and Sample B. These results are summarized via the squared multiple correlations presented in Table 2. In the columns headed "Ridge" are squared multiple correlations calculated as part of the matrix manipulations during the ridge regression analysis. The columns headed "Cross" contain squared multiple correlations calculated by predicting standardized scores for each district and dividing the variance of predicted scores by the variance of the actual standardized scores. In all cases the squared multiple correlations decreased as the constant added to the diagonal increased.

When the betas from Sample A were applied to Sample B for prediction, the squared multiple correlations from the cross-validation were less than those from the ridge regression for all of Example 1 and Example 2 and for most of Example 3. The one exception is in Example 3 when k = 0 so that classical multiple regression

betas are utilized. In this case the cross-validation has a higher squared multiple correlation.

When the betas from Sample B were applied to Sample A for prediction, the squared multiple correlations from the cross-validation were less than those from the ridge regression for all of Example 1 and Example 2 and for much of Example 3. For Example 3 the cross-validation yielded higher squared multiple correlations when utilizing the values of the constant from .00 to .35. Note the very sizeable difference from .394 to .532 when cross-validating the betas from classical multiple regression. The reason for this finding when the constant is small and only in this one example is not understood at this time.

When the betas from a sample are applied to the standardized scores in the same sample, it should be possible to verify the squared multiple correlations calculated from the ridge regression. When the constant is zero, the two estimates are equal. As the constant increases so does the difference in the two squared multiple correlations, always favoring the analytical estimate.

The implications of the findings presented here are that researchers using highly correlated independent variables should be cautious when interpreting the regression coefficients resulting from multiple linear regression. In such situations ridge regression could be utilized as an exploratory tool to search for regression coefficients which may yield more meaningful interpretations. However, given the results of the cross-validation, it would seem that further exploratory work concerning the nature of ridge regression needs to be carried out before it is widely applied.

Table 1. Standardized Regression Coefficient From Ridge Regression

SAMPLE A	Constant Added	Multiple Correlation Squared	Word Relation-Ships	Reading	English	Math	Instr. Staff per 1000 Students	Teachers per 1000 Students	Tchrs Years Exp.	Teachers Salary	Percent Teachers w/ Masters
Example 1	.00	.240	.237	.257	-.069	.079					
(Grade 4 Achievement	.50	.204	.139	.144	.071	.090					
	.80	.189	.123	.127	.077	.088					
Example 2	.00	.158					-.675	.797	.197	-.270	.356
(Teacher Variables)	.50	.073					.012	.101	.155	-.096	.115
	.80	.060					.022	.081	.128	-.064	.087
Example 3	.00	.394	.413	.274	-.094	-.096	-.720	.810	.255	-.292	.205
(Grade 4 Achievement	.50	.268	.151	.150	.076	.066	.004	.095	.160	-.125	.048
+ Teacher Variables)	.80	.241	.129	.129	.081	.074	.016	.077	.128	-.090	.035
Correlation with Grade 7 Achievement			.477	.476	.430	.429	.118	.169	.230	-.026	.142
SAMPLE B											
Example 1	.00	.381	.038	.159	.039	.409					
(Grade 4 Achievement)	.50	.325	.110	.125	.121	.205					
	.80	.302	.110	.120	.118	.175					
Example 2	.00	.050					-.212	.237	.160	-.156	.142
(Teacher Variables)	.50	.026					-.009	.047	.110	-.069	.041
	.80	.021					-.001	.037	.091	-.051	.029
Example 3	.00	.417	.136	.175	-.041	.402	.004	-.135	.095	-.231	.130
(Grade 4 Achievement	.50	.342	.124	.127	.115	.203	-.032	-.033	.066	-.085	.031
and Teacher Variables)	.80	.315	.112	.121	.115	.174	-.027	-.022	.057	-.063	.019
Correlation with Grade Achievement			.555	.566	.569	.605	.039	.078	.159	-.052	.043

Table 2. Squared Multiple Correlations from Ridge Equations and from Cross-Validation

		Example 1		Example 2		Example 3	
	Constant	Ridge	Cross	Ridge	Cross	Ridge	Cross
A applied to B	.00	.381	.378	.050	.042	.417	.437
	.25	.348	.320	.033	.021	.371	.34[illegible]
	.50	.325	.280	.026	.014	.342	.293
	.75	.305	.249	.022	.010	.319	.256
	1.00	.288	.222	.019	.008	.300	.227
B applied to A	.00	.240	.238	.158	.211	.394	.532
	.25	.219	.203	.092	.076	.302	.312
	.50	.204	.179	.073	.051	.268	.255
	.75	.192	.158	.062	.038	.245	.217
	1.00	.181	.142	.054	.030	.226	.189
A applied to A	.00	.240	.240	.158	.158	.394	.394
	.25	.219	.202	.092	.066	.302	.258
	.50	.204	.177	.073	.045	.268	.213
	.75	.192	.157	.062	.033	.245	.184
	1.00	.181	.140	.054	.026	.226	.161
B applied to B	.00	.381	.381	.050	.050	.417	.417
	.25	.348	.322	.033	.023	.371	.337
	.50	.325	.283	.026	.016	.342	.292
	.75	.305	.251	.022	.011	.319	.257
	1.00	.288	.224	.019	.009	.300	.228

Figure 1. Sample A Fourth Grade Tests Predicting Seventh Grade Composite Achievement: Standardized Regression Coefficients by Constant Added to the Diagonal.

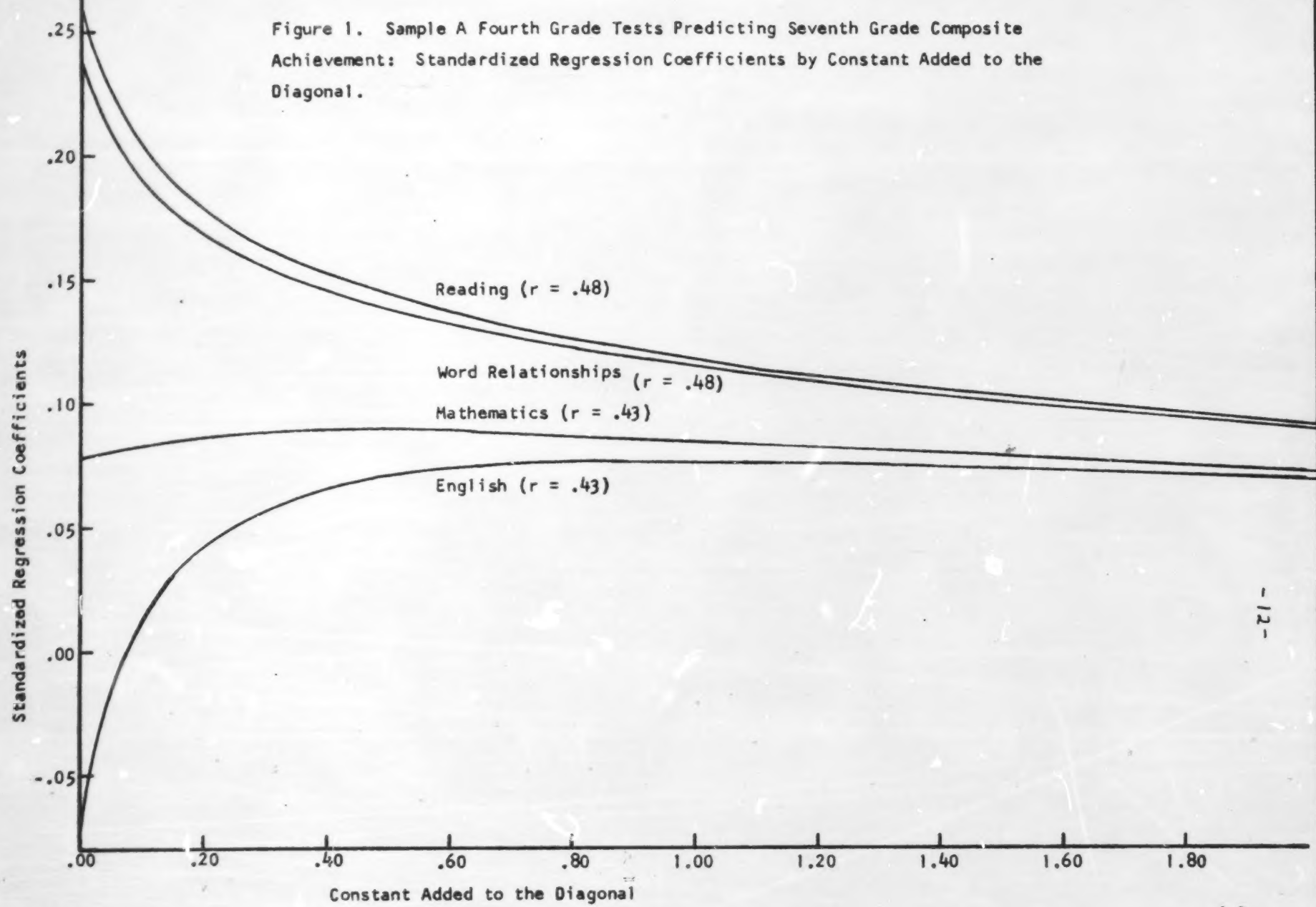


Figure 2. Sample B Fourth Grade Tests Predicting Seventh Grade Composite Achievement: Standardized Regression Coefficients by Constant Added to the Diagonal.

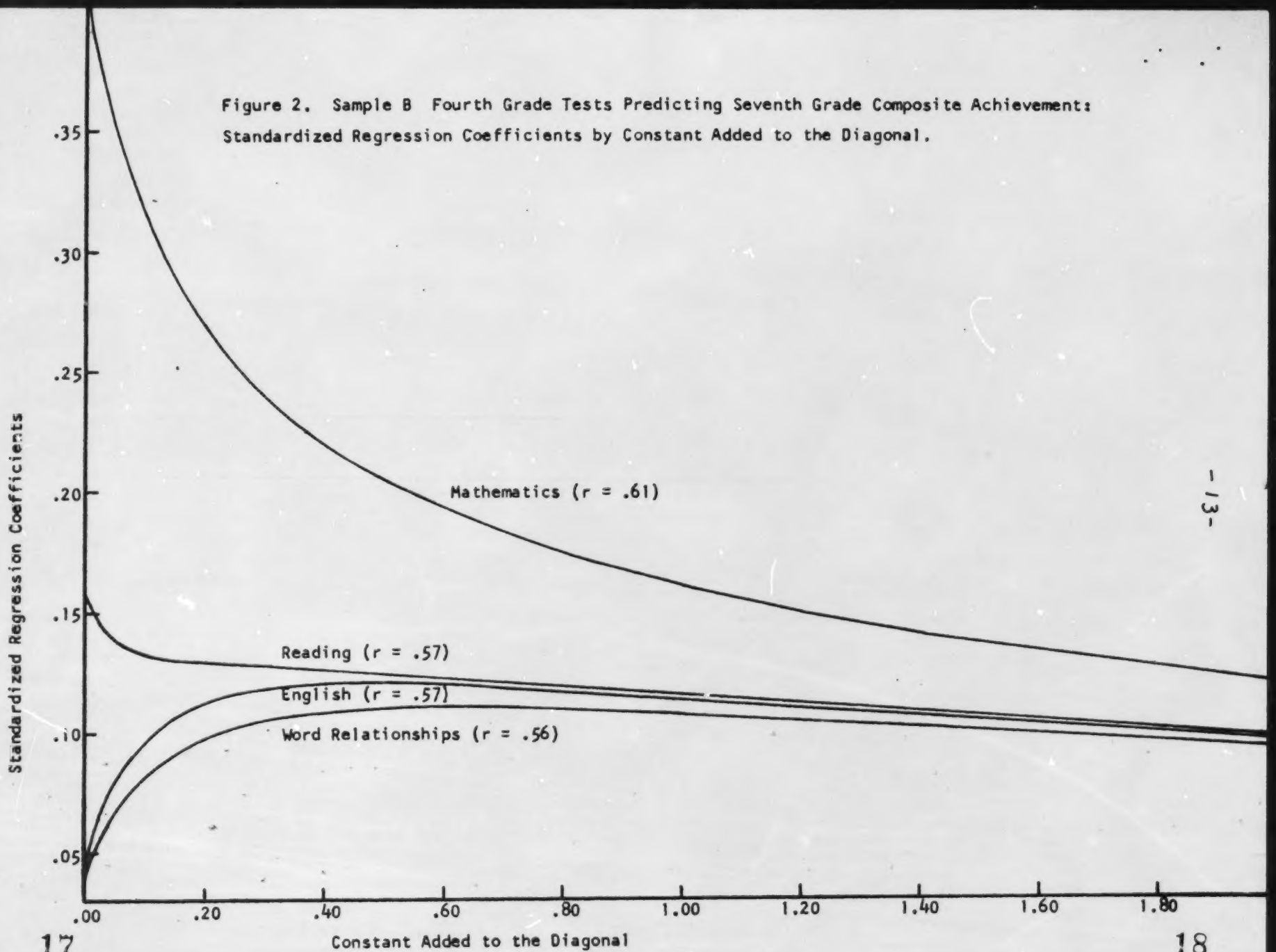


Figure 3. Sample A Teacher Variables Predicting Seventh Grade Achievement: Standardized Regression Coefficients by Constant Added to the Diagonal.

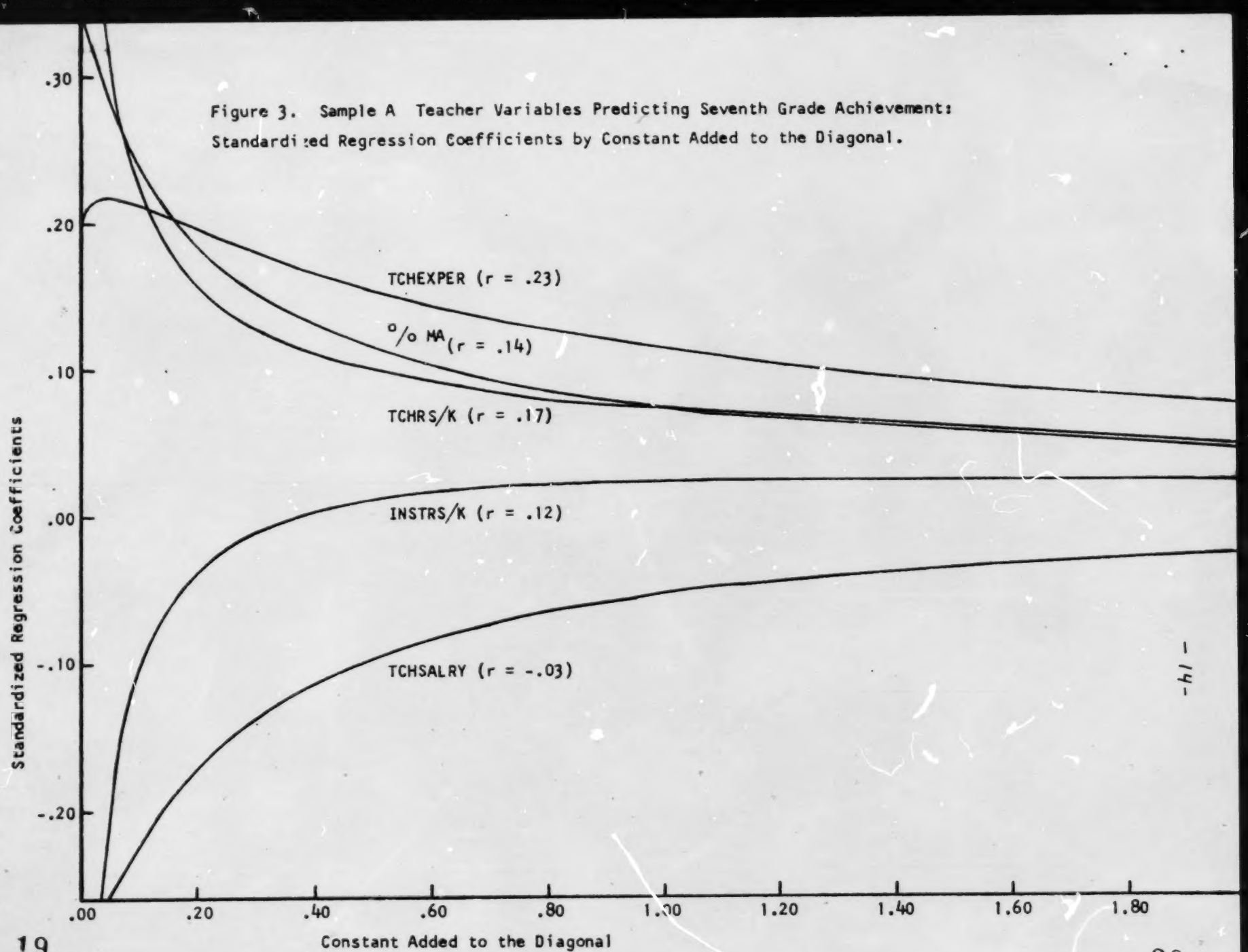


Figure 4. Sample B Teacher Variables Predicting Seventh Grade Achievement: Standardized Regression Coefficients by Constant Added to the Diagonal.

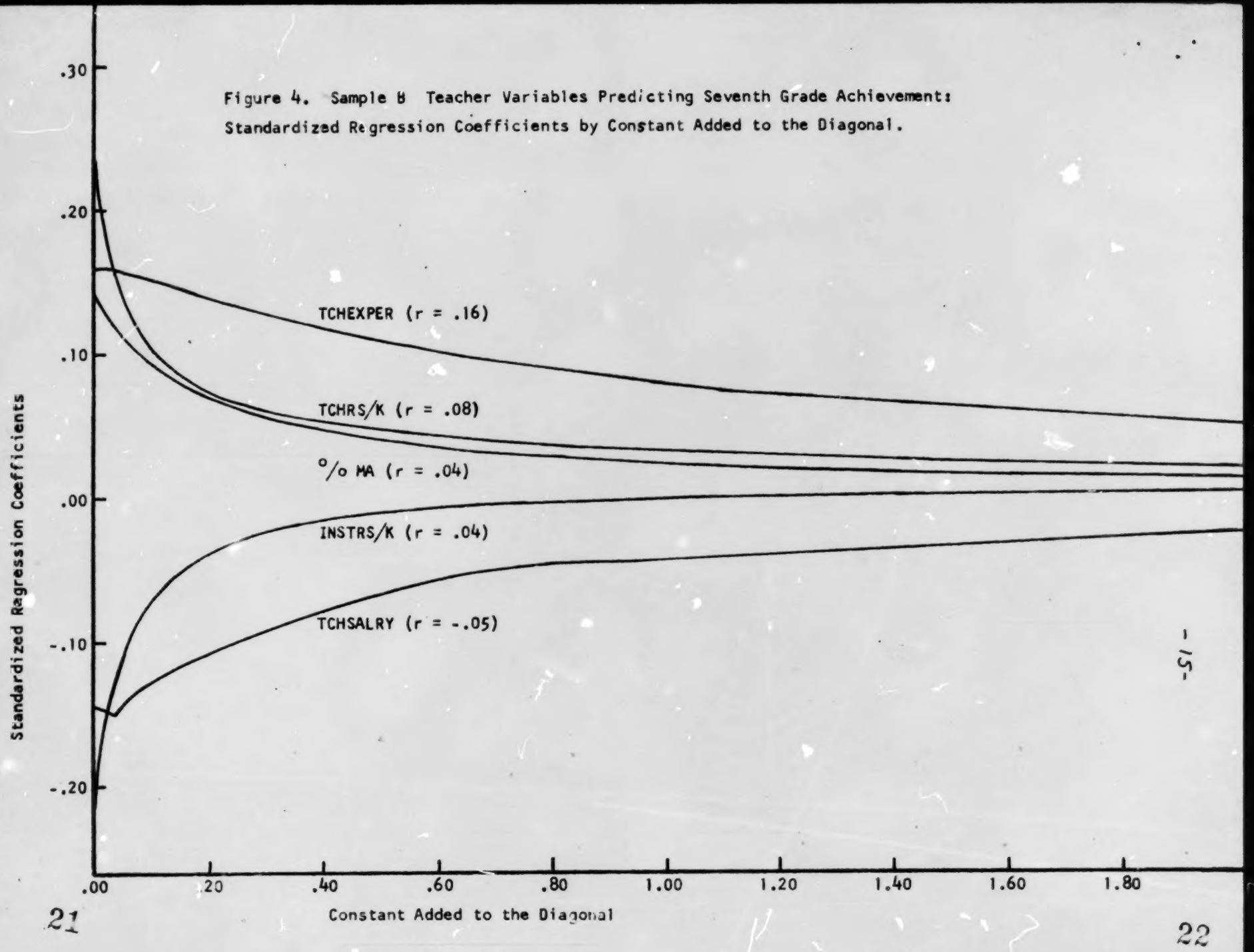


Figure 5. Sample A Teacher Variables and Fourth Grade Tests Predicting Seventh Grade Achievement: Standardized Regression Coefficients by Constant Added to the Diagonal.

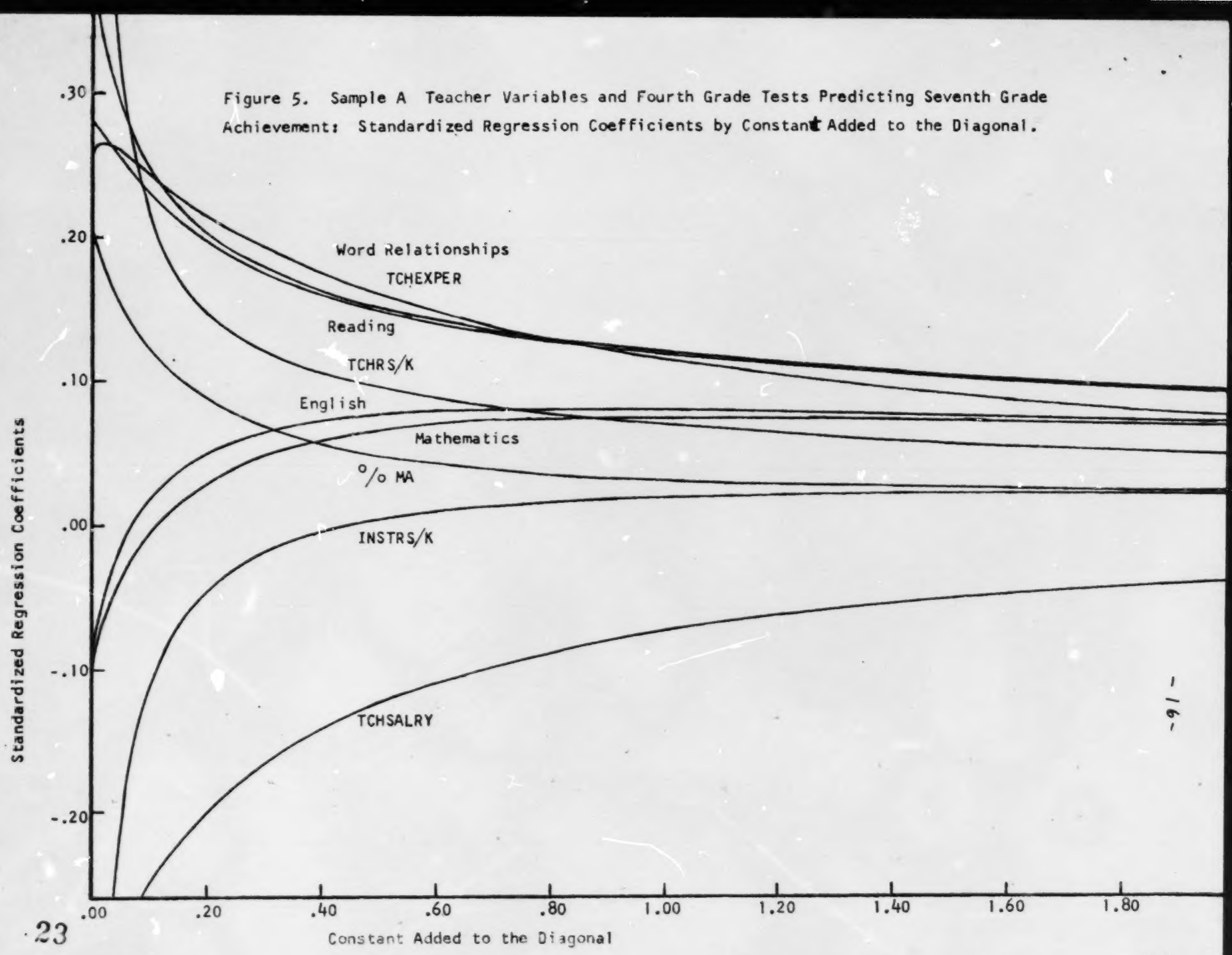
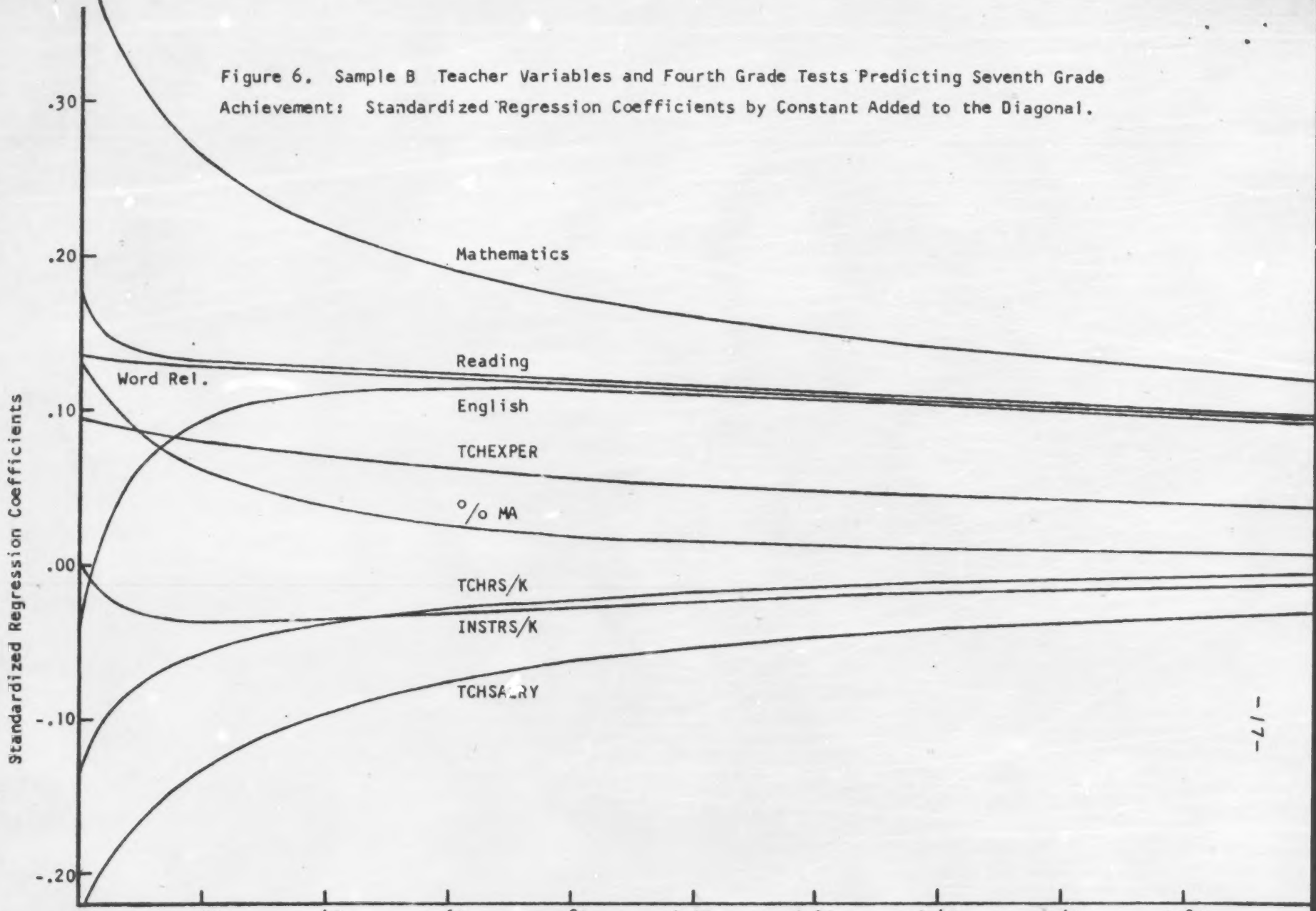


Figure 6. Sample B Teacher Variables and Fourth Grade Tests Predicting Seventh Grade Achievement: Standardized Regression Coefficients by Constant Added to the Diagonal.



REFERENCES

Coleman, James S., et al., Equality of Educational Opportunity, Washington D.C.: Office of Education, U. S. Department of Health, Education and Welfare, 1966.

Draper, N. R., "Ridge Analysis"of Response Surfaces. Technometrics, 1963, 5., 469-479.

Hoerl, Arthur E., Application of Ridge Analysis to Regression Problems. Chemical Engineering Progress, 1962, 58, 54-59.

Hoerl, Arthur E. and Kennard, R. W., Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, 1970a, 12, 55-67.

Hoerl, Arthur E. and Kennard, R. W., Ridge Regression: Applications to Nonorthogonal Problems. Technometrics, 1970b, 12, 69-82.

Hoerl, Arthur E., Kennard, R. W., and Baldwin, K. F., Ridge Regression: Some Simulations. Communication in Statistics, 1975, 4, 105-123.

Husen, Torsten, International Study of Achievement in Mathematics: A Comparison of Twelve Countries. New York: John Wiley + Sons, 1967.

Madaus, G. F., Kellaghan, T. and Rakow, E. A., A Study of the Sensitivity of Measures of School Effectiveness. Report submitted to the Carnegie Corporation, New York, December, 1975.

Marquardt, D. W. and Snee, R. D., Ridge Regression in Practice. The American Statistician, 1975, 29, 3-19.

Newton, R. G. and Spurrell, D. J., A Development of Multiple Regression for the Analysis of Routine Data, Applied Statistics, 1967, 16, 51-64.

Newton, R. G. and Spurrell, D. J., Examples of the Use of Elements for Clarifying Regression Analysis. Applied Statistics, 1967, 16, 165-172.

Peaker, G. F., An Empirical Study of Education in Twenty-One Countries: A Technical Report. New York: John Wiley + Sons, 1975.

Williams, S. H., An Application of Ridge Regression to the Prediction of Enrollment in Higher Education. Paper presented at Mid-South Educational Research Association Annual Convention, November, 1975.